University of Zurich

Socioeconomic Institute
Sozialökonomisches Institut

Working Paper No. 0918

Partial Identification of Discrete Counterfactual Distributions with Sequential Update of Information

Stefan Boes

December 2009

Socioeconomic Institute
University of Zurich

Working Paper No. 0918

Partial Identification of Discrete Counterfactual Distributions with Sequential Update of Information

December 2009

Author's address: Stefan Boes
E-mail: boes@sts.uzh.ch

Publisher Sozialökonomisches Institut
Bibliothek (Working Paper)
Rämistrasse 71
CH-8006 Zürich
Phone: +41-44-634 21 37
Fax: +41-44-634 49 82
URL: www.soi.uzh.ch
E-mail: soilib@soi.uzh.ch

Partial Identification of Discrete Counterfactual Distributions With Sequential Update of Information*

Stefan Boes†

December 14, 2009

Abstract

The credibility of standard instrumental variables assumptions is often under dispute. This paper imposes weak monotonicity in order to gain information on counterfactual outcomes, but avoids independence or exclusion restrictions. The outcome process is assumed to be sequentially ordered, building up and depending on the information level of agents. The potential outcome distribution is assumed to weakly increase (or decrease) with the instrument, conditional on the continuation up to a certain stage. As a general result, the counterfactual distributions can only be bounded, but the derived bounds are informative compared to the no-assumptions bounds thus justifying the instrumental variables terminology. The construction of bounds is illustrated in two data examples.

JEL Classification: C14, C25, C35

Keywords: Nonparametric bounds, treatment effects, endogeneity, binary choice, monotone instrumental variables, policy evaluation.

*Acknowledgments: Financial support was generously provided through Swiss NSF grant #PBZHP1-124191. I thank Karim Chalak, Gary Chamberlain, Rustam Ibragimov, Guido Imbens, Herman van Dijk, and seminar participants at Harvard for valuable comments.

†Address for correspondence: University of Zurich, Socioeconomic Institute, Zuerichbergstrasse 14, CH-8032 Zurich, Switzerland, phone: +41 44 634 2301, email: boes@sts.uzh.ch.; Harvard University, Institute for Quantitative Social Science, 1730 Cambridge Street, Cambridge, MA 02138, USA, phone: +1 617 649 9632, email: sboes@iq.harvard.edu.

1 Introduction

The empirical evaluation of interventions is a major task of economic policy-making. One might be interested, for example, in the effect of a supplementation policy on the education outcomes of children in disadvantaged families (e.g., Brooks-Gunn 2003), or in the effect of schooling on fertility (e.g., Sander 1992). The potential outcomes framework (Rubin 1974) has become a popular tool to assess such interventions. More formally, let $Y(s)$ denote the potential outcome in state $s \in \mathcal{S}$, and let $P[Y(s)]$ denote the distribution of outcomes that would occur if everybody in the population was observed in state s. The study of policy effects compares the potential outcome distributions in different states, invoking assumptions on the nature of the data-generating process and the information available to agents.

The central issues in this framework are twofold. First, treatment states are generally subject to choice, e.g., a family can decide whether or not to accept a financial aid. Second, the realized outcome of each individual is only observed in the actual state, the outcomes that would occur in alternative states are logically unobserved. As a result, the commonly defined policy effects are fundamentally unidentified (Manski 1994, 2000, 2007, Heckman and Vytlacil 2007a, 2007b, among others). Possible solutions require assumptions on the data-generating process, either by specifying a parametric model on the potential outcomes (e.g., Haavelmo 1943, Roy 1951, Heckman and Honoré 1990, Heckman 2001, Heckman and Vytlacil 2005), or by restricting the attention to some features of the population, such as the mean or the median, and using the observed data combined with some less restrictive assumptions to infer these parameters (e.g., Imbens and Angrist 1995, Angrist *et al.* 1996). As a third alternative, one may impose bounds on the policy effects of interest under weak assumptions on the data structure (e.g., Manski 1990, 1997, Manski and Pepper 2000, Shaikh and Vytlacil 2005).

This paper follows the latter strategy. It focuses on outcomes that are measured on a discrete ordinal scale with the agent's information set being generated by a sequential mechanism. The sequential structure is motivated by an economic model of transitions, where certain responses are only observed given positive outcomes in the antecedent responses (Amemiya

1975, 1985). In this model, one may introduce latent variables that represent marginal returns, based on the net lifetime reward from stopping at some stage (Cunha *et al.* 2007, Carneiro *et al.* 2003). The optimal stopping time is then characterized by that stage when marginal returns are just above zero, and the marginal returns for all adjacent stages are negative. Examples include schooling outcomes (Steele *et al.* 2009), the number of children (Zhang 1994), the number of unemployment spells (Kahn and Morimune 1979), and the labor force participation of women (Heckman and Willis 1977).

In order to gain information on the counterfactual outcome distributions, weak assumptions are imposed on the sequential decision process. In a first step, I will explore monotone instrumental variables (MIV) assumptions in the manner of Manski and Pepper (2000, 2009) but extended to incorporate stochastic dominance. In a second step, I will relax these assumptions to allow for partially monotone instrumental variables (PMIV), taking into account the information set in each stage. As a general result, the counterfactual distributions can only be bounded, with the width of bounds depending on the strength of the underlying assumptions. To the best of my knowledge, the PMIV approach is new in the literature.

The analysis is somewhat related to the dynamic treatment literature due to the sequential ordering of outcomes (e.g., Robins 1989, 1997, Lechner and Miquel 2009, Gill and Robins 2001, Lechner 2008, 2009, Abbring and Van Den Berg 2003, Navarro and Heckman 2007, Abbring and Heckman 2007). However, it differs from that literature in the nature of the data-generating process. Here, potential outcomes arise from a sequential model with available information depending on previous experience (or outcomes). The model is static in the sense that the intervention only occurs and affects choices at a particular point in time. The approach also differs from Boes (2007) who imposes a multiple threshold crossing model that generates the ordered potential outcomes.

The remainder of the paper is organized as follows. Section 2 provides a heuristic outline of the analysis. Section 3 develops the concept of PMIV and derives bounds on the counterfactual distributions. Section 4 illustrates the construction of bounds using a simulated and a real data example. The empirical application considers how the number of children born to a woman varies with realized schooling, under the assumption that fertility decisions are based on both current circumstances and experiences with previously born children (Zhang 1994). This example will appear throughout the text in order to motivate the MIV and PMIV assumptions. Section 5 summarizes the results and concludes.

2 Heuristic Derivation of Bounds

A simple example with $J = 3$ ordered responses illustrates the strategy of the paper. Let $S \in \mathcal{S}$ denote the state in which an individual is observed, and let $Y = Y(S)$ denote the realized outcome with $Y \in \mathcal{Y} = \{1, ..., J\}$. Outcomes are ordered such that $1 < ... < J$, although the distance between outcomes does not necessarily have an interpretation. The observed data are the triple (Y, S, X) of realized outcomes Y, realized states S, and covariates $X \in \mathcal{X}$. The problem is to learn the distribution of potential outcomes $P[Y(s)]\ \forall s \in \mathcal{S}$, perhaps conditional on X, from the observed distribution $P(Y, S, X)$. Consider the distribution of outcomes in state s

$$P[Y(s)] = P[Y|S = s]\, P(S = s) + P[Y(s)|S \neq s]\, P(S \neq s), \tag{1}$$

which follows by the law of total probability and the observation rule for Y. The potential outcome $Y(s)$ is not observed for individuals not in s, and thus the (counterfactual) distribution $P[Y(s)|S \neq s]$ is not identified from the observed data. All other quantities are identified. By definition of a probability, it must hold that $P[Y(s)|S \neq s] \in [0, 1]$ which yields an identification region for the distribution of $Y(s)$ (Manski 1995, 2003).

Now rewrite the distribution in terms of conditional transition probabilities, formally $P[Y(s) = y] \equiv P[Y(s) = y|Y(s) \geq y]\, P[Y(s) \geq y]$, $\forall y \in \mathcal{Y}$. All the assumptions on the data-generating process will be stated with respect to the transition probabilities $P[Y(s) = y|Y(s) \geq y]$, reflecting the information that agents have when deciding upon the outcome of interest. Consider the counterfactual distribution with $J = 3$ outcomes:

$$\begin{aligned} P[Y(s) = 1|S \neq s] &= P[Y(s) = 1|Y(s) \geq 1, S \neq s] \\ P[Y(s) = 2|S \neq s] &= P[Y(s) = 2|Y(s) \geq 2, S \neq s] \\ &\quad \cdot\{1 - P[Y(s) = 1|Y(s) \geq 1, S \neq s]\} \\ P[Y(s) = 3|S \neq s] &= \prod_{j=1}^{2}\{1 - P(Y(s) = j|Y(s) \geq j, S \neq s]\}. \end{aligned} \tag{2}$$

Without imposing further assumptions on the data-generating process, each of the conditional transition probabilities must lie within the unit interval, and thus

$$P[Y(s) = y|S \neq s] \in [0, 1] \quad \forall y \in \{1, 2, 3\}. \tag{3}$$

These are the no-assumptions bounds mentioned above. Now assume that the conditional transition probability for the lowest category can be restricted due to some additional information, but there are no restrictions on the transition from the second to the third category (given the information from the first). This implies

$$\begin{aligned} P[Y(s) = 1|Y(s) \geq 1, S \neq s] &\in [\alpha_l, \alpha_u] \\ P[Y(s) = 2|Y(s) \geq 2, S \neq s] &\in [0, 1]. \end{aligned} \tag{4}$$

In this case, bounds on the unidentified probability distribution can be derived as follows:

$$\begin{aligned} P[Y(s) = 1|S \neq s] &\in [\alpha_l, \alpha_u] \\ P[Y(s) = 2|S \neq s] &\in [0, 1 - \alpha_l] \\ P[Y(s) = 3|S \neq s] &\in [0, 1 - \alpha_l]. \end{aligned} \tag{5}$$

The assumption imposed in the left part of the outcome distribution, in terms of the conditional transition probability, passes through to all subsequent (unconditional) probabilities and thus is informative also for the right part of the outcome distribution. Analogous arguments hold if additional restrictions are imposed on $P[Y(s) = 2|Y(s) \geq 2, S \neq s]$. The next two sections discuss weak (but credible) assumptions on the data-generating process that can be invoked to bound the conditional transition probabilities.

3 (Partially) Monotone Instrumental Variables

Manski and Pepper (2000, 2009) introduce monotone instrumental variables (MIV) in order to weaken conventional instrumental variables (IV) assumptions. IV assumptions typically impose mean or full independence of the response variable and the instrument. MIV assumptions replace the equality of mean outcomes (or outcome distributions) of standard IV methods by a weak inequality conditional on the instrument. This section further develops the idea in the context of an ordered discrete outcome model with sequential update of information. The notation follows Manski and Pepper (2000, 2009).

Let $X = (W, V)$ and $\mathcal{X} = \mathcal{W} \times \mathcal{V}$. Each value of (W, V) characterizes a subpopulation of individuals. Standard IV methods state that V is an instrument if the distribution of outcomes for each $s \in \mathcal{S}$ conditional on W does not change with different values of V. Formally, the IV assumptions may relate to the entire distribution or the conditional transition probabilities:

Assumption (IV1)***.*** *Covariate V is an IV in the sense of conditional independence in the distribution of Y if, for each $y \in \mathcal{Y}$, $s \in \mathcal{S}$, $W \in \mathcal{W}$, and all $u_1, u_2 \in \mathcal{V} \times \mathcal{V}$,*

$$P[Y(s) = y|W, V = u_1] = P[Y(s) = y|W, V = u_2]. \tag{6}$$

Assumption (IV2)***.*** *Covariate V is an IV in the sense of conditional independence in the transition probabilities if, for each $y \in \mathcal{Y}$, $s \in \mathcal{S}$, $W \in \mathcal{W}$, and all $u_1, u_2 \in \mathcal{V} \times \mathcal{V}$,*

$$P[Y(s) = y|Y(s) \geq y, W, V = u_1] = P[Y(s) = y|Y(s) \geq y, W, V = u_2]. \tag{7}$$

Now assume that $\mathcal{V}$ is an ordered set. A distributional equivalent to the MIV assumption of Manski and Pepper (2000) would be that the distribution of $Y(s)$ for all individuals with a specified W and value u_1 of V weakly dominates the distribution of $Y(s)$ for individuals with the same W but with $u_2 \geq u_1$, $(u_2, u_1) \in \mathcal{V}$. Formally:

Assumption (MIV1)***.*** *Covariate V is an IV in the sense of weak stochastic dominance if, for each $y \in \mathcal{Y}$, $s \in \mathcal{S}$, $W \in \mathcal{W}$, and all $u_1, u_2 \in \mathcal{V} \times \mathcal{V}$ with $u_1 \leq u_2$,*

$$P[Y(s) \leq y|W, V = u_1] \geq P[Y(s) \leq y|W, V = u_2]. \tag{8}$$

An assumption of this type is explored in greater detail in Boes (2009). In the context of this paper and the sequential mechanism generating the outcomes, it is more reasonable to specify the MIV assumption in terms of transition probabilities.

Assumption (MIV2)***.*** *Covariate V is an IV in the sense of weak stochastic dominance in the transition probabilities if, for each $y \in \mathcal{Y}$, $s \in \mathcal{S}$, $W \in \mathcal{W}$, and all $u_1, u_2 \in \mathcal{V} \times \mathcal{V}$ with $u_1 \leq u_2$,*

$$P[Y(s) = y | Y(s) \geq y, W, V = u_1] \geq P[Y(s) = y | Y(s) \geq y, W, V = u_2]. \tag{9}$$

In order to illustrate how the assumptions IV1 and MIV1 differ, consider the following example with outcome of interest the number of children born to a woman, S the woman's schooling, and V measured marital attractiveness. For V to serve as an instrument in the sense of IV1, we must assume that the distribution of children for schooling level s is the same for women with a high attractiveness than for women with a low attractiveness. Assumption MIV1 states that less attractive women tend to have a higher probability of getting few children, and more attractive women tend to have a higher probability of getting many children. If attractivenss is a complement to child-rearing, then the latter assumption is consistent with economic theory (e.g., Boulier and Rosenzweig 1984), assumption IV1 is not.

Likewise, assumption IV2 claims that the probability of having y children, given at least y children, is the same for those women who are less attractive and those women who are more attractive in the marital market. Assumption MIV2 states that this probability is weakly higher for less attractive women than for more attractive women. In the language of discrete hazard models (Lancaster 1990), the probability of a drop-out at a certain stage (where stage is a particular number of children), conditional on survival up to this stage (i.e., having at least that many children), is weakly higher for the less attractive women than for the more attractive women.

The focus here is on the implications of assumption MIV2. In order to simplify the notation, I will drop W, but all expressions are implicitly assumed conditional on W. MIV2 implies for any $u \in \mathcal{V}$ with $u_1 \leq u \leq u_2$ that

$$\begin{aligned} P[Y(s) = y|Y(s) \geq y, V = u_2] \;&\leq\; P[Y(s) = y|Y(s) \geq y, V = u] \qquad (10) \\ \leq\;& P[Y(s) = y|Y(s) \geq y, V = u_1], \quad \forall y \in \mathcal{Y}. \end{aligned}$$

Thus, a lower bound of $P[Y(s) = y|Y(s) \geq y, V = u_2]$ is also a lower bound of $P[Y(s) = y|Y(s) \geq y, V = u]$, and an upper bound of $P[Y(s) = y|Y(s) \geq y, V = u_1]$ is also an upper bound of $P[Y(s) = y|Y(s) \geq y, V = u]$. This must hold for all values $u_1 \leq u$ and must hold for all values $u_2 \geq u$. Exploiting the full support of V, one can impose bounds on the distribution of potential outcomes:

Proposition 1. *Let assumption MIV2 hold. Then for all $y \in \mathcal{Y}$ and any $u \in \mathcal{V}$,*

$$\begin{aligned} &\sup_{u_2 \geq u} \{P[Y = y|Y \geq y, S = s, V = u_2]P(S = s|V = u_2)\} \qquad (11) \\ &\leq\; P[Y(s) = y|Y(s) \geq y, V = u] \;\leq \\ &\inf_{u_1 \leq u} \{P[Y = y|Y \geq y, S = s, V = u_1]P(S = s|V = u_1) + P(S \neq s|V = u_1)\}. \end{aligned}$$

In the absence of other information, these bounds are sharp.

Bounds on $P[Y(s) = y|Y(s) \geq y]$ can be obtained by integrating out over the support of V (subject to conditions of measurability). Proposition 1 is similar to the one in Manski and Pepper (2000), taking into account that probabilities are equivalent to expectations of value indicators of Y. It differs from the proposition in Manski and Pepper in the conditioning argument, reflecting the sequential update of information in the data-generating process.

It is exactly the sequential structure that can be explored to further relax the monotone instrumental variables assumptions. More specifically, there is no need to impose weak dominance throughout the entire support of Y, and one might relax assumption MIV2 to a partially monotone instrumental variables (PMIV) assumption. Two versions of the PMIV assumptions can be stated. These are:

Assumption (PMIV1)***.*** *Covariate V is an IV in the sense of partially weak stochastic dominance with monotone switch in the transition probabilities if, for each $y \in \mathcal{Y}$, $s \in \mathcal{S}$, $W \in \mathcal{W}$,* *and all* $u_1, u_2 \in \mathcal{V} \times \mathcal{V}$ *with* $u_1 \leq u_2$*,*

$$\begin{aligned} P[Y(s) = y|Y(s) \geq y, V = u_1] &\geq P[Y(s) = y|Y(s) \geq y, V = u_2], \ y \leq t \\ P[Y(s) = y|Y(s) \geq y, V = u_1] &\leq P[Y(s) = y|Y(s) \geq y, V = u_2], \ y > t \end{aligned} \tag{12}$$

where $t \in \mathcal{Y}$ *is a fixed threshold value.*

Assumption (PMIV2)***.*** *Covariate* V *is an IV in the sense of partially weak stochastic dominance in the transition probabilities if, for* $y, y' \in \mathcal{Y} \times \mathcal{Y}$ *and* $y \neq y'$*, and for each* $s \in \mathcal{S}$*,* $W \in \mathcal{W}$*, and all* $u_1, u_2 \in \mathcal{V} \times \mathcal{V}$ *with* $u_1 \leq u_2$*,*

$$\begin{aligned} P[Y(s) = y|Y(s) \geq y, V = u_1] &\geq P[Y(s) = y|Y(s) \geq y, V = u_2], \\ P[Y(s) = y'|Y(s) \geq y', V = u_1] &\leq P[Y(s) = y'|Y(s) \geq y', V = u_2]. \end{aligned} \tag{13}$$

In order to illustrate assumptions PMIV1 and PMIV2, consider the schooling and fertility example again. Assumption PMIV1 states that the probability of getting a low number of children ($y \leq t$, for some fixed t), given the information from previously born children (if any), is weakly higher for the less attractive women ($V = u_1$) compared to the more attractive women ($V = u_2$). The drop-out rate given a high number of children ($y > t$), however, may be higher for the more attractive women than the less attractive women, e.g., due to the counteraction of income and substitution effects (again see Boulier and Rosenzweig 1984). If there does not exist a unique threshold t at which the order of weak inequalities changes (e.g., due to the characteristics of the population), then there might be some $y \in \mathcal{Y}$ where the drop-out rate is higher for the more attractive women, and some other $y' \in \mathcal{Y}$ where the drop-out rate is higher for the less attractive women, without particular order. The latter is consistent with assumption PMIV2.

Since PMIV1 is a special case of PMIV2, only the implications of the latter will be discussed (with the implications of the former being immediate). Given assumption PMIV2 it must hold that for any $u \in \mathcal{V}$ with $u_1 \leq u \leq u_2$ and $y, y' \in \mathcal{Y} \times \mathcal{Y}$, $y \neq y'$ as specified by the assumption,

$$P[Y(s) = y|Y(s) \geq y, V = u_1] \ \geq \ P[Y(s) = y|Y(s) \geq y, V = u] \tag{14}$$

$$
\begin{aligned}
&\geq \; P[Y(s) = y|Y(s) \geq y, V = u_2], \\
P[Y(s) = y'|Y(s) \geq y', V = u_1] \; &\leq \; P[Y(s) = y'|Y(s) \geq y', V = u] \qquad (15) \\
&\leq \; P[Y(s) = y'|Y(s) \geq y', V = u_2].
\end{aligned}
$$

Thus, it follows from (14) that a lower bound of $P[Y(s) = y|Y(s) \geq y, V = u_2]$ is also a lower bound of $P[Y(s) = y|Y(s) \geq y, V = u]$, and an upper bound of $P[Y(s) = y|Y(s) \geq y, V = u_1]$ is also an upper bound of $P[Y(s) = y|Y(s) \geq y, V = u]$. This must hold for all $u_1 \leq u$ and all $u_2 \geq u$. Similar arguments, but with opposite signs, hold for the transition probabilities of category $y' \neq y$ as captured by the second part of assumption PMIV2 and (15). Exploiting the full support of V yields:

Proposition 2. *Let assumption PMIV2 hold. Then for any $u \in \mathcal{V}$,*

$$
\begin{aligned}
&\sup_{u_2 \geq u} \{P[Y = y|Y \geq y, S = s, V = u_2]P(S = s|V = u_2)\} \qquad (16) \\
&\quad \leq \; P[Y(s) = y|Y(s) \geq y, V = u] \; \leq \\
&\inf_{u_1 \leq u} \{P[Y = y|Y \geq y, S = s, V = u_1]P(S = s|V = u_1) + P(S \neq s|V = u_1)\},
\end{aligned}
$$

and

$$
\begin{aligned}
&\sup_{u_1 \leq u} \{P[Y = y'|Y \geq y', S = s, V = u_1]P(S = s|V = u_1)\} \qquad (17) \\
&\quad \leq \; P[Y(s) = y'|Y(s) \geq y', V = u] \; \leq \\
&\inf_{u_2 \geq u} \{P[Y = y'|Y \geq y', S = s, V = u_2]P(S = s|V = u_2) + P(S \neq s|V = u_2)\}.
\end{aligned}
$$

In the absence of other information, these bounds are sharp.

Proposition 2 consists of two parts that reflect the different information on the sequential decision process, depending on the outcome level itself. (16) is derived from the first part of PMIV2 and the inequalities in (14). The bounds in (17) are derived from the second part of PMIV2 and the inequalities in (15). Since no other assumptions on the data-generating process are imposed than PMIV2, it is difficult to draw general conclusions regarding the properties of the bounds. Nevertheless, some remarks can be made. First, the bounds in (16) are not informative if the lower and upper no-assumptions bounds on $P[Y(s) = y|Y(s) \geq y, V = u]$ weakly decrease with u, in which case they coincide with the no-assumptions bounds. Second, the bounds coincide with the standard IV bounds (under IV2) if the no-assumptions lower and upper bounds on $P[Y(s) = y|Y(s) \geq y, V = u]$ weakly increase with u. The opposite holds for the second part of proposition 2, i.e., the stated bounds coincide with the no-assumptions bounds (with the IV2 bounds) if the lower and upper no-assumptions bounds on $P[Y(s) = y|Y(s) \geq y, V = u]$ weakly increase (decrease) with u.

Monotone treatment selection

A special case is obtained if V is the treatment indicator itself. Assumption IV1 then states that the potential outcome distribution is the same irrespective of the treatment status, formally $P[Y(s)|S = u_1] = P[Y(s)|S = u_2]$ for all $u_1, u_2 \in \mathcal{S} \times \mathcal{S}$. This the exogenous treatment selection assumption known from the literature (e.g., Manski and Pepper 2000). Note that further covariates W are kept implicit in the conditioning part. Now suppose that $\mathcal{S}$ is an ordered set and $V = S$. In this case assumptions MIV1, MIV2, PMIV1, and PMIV2 relax the exogenous treatment selection assumption to monotone treatment selection (MTS) assumptions (cf. Manski and Pepper 2000), or partially monotone treatment selection (PMTS) assumptions. For example, assumption MIV2 states that the drop-out rate for those with a low value of S is weakly higher than the drop-out rate for those with a high value of S. Likewise, assumption PMIV2 states that the sign of the weak inequality between drop-out rates depends on the outcome level.

The following corollary summarizes the results under the partially monotone treatment selection version of assumption PMIV2:

Corollary 1. *Let $\mathcal{S}$ be an ordered set and let assumption PMIV2 hold for $V = S$. Then, for any $u \in \mathcal{S}$ the bounds in Proposition 2 reduce to*

$$u < s \quad \Rightarrow \quad P(Y = y|Y \geq y, S = s) \leq P[Y(s) = y|Y(s) \geq y, S = u] \leq 1 \tag{18}$$

$$u = s \quad \Rightarrow \quad P[Y(s) = y|Y(s) \geq y, S = u] = P(Y = y|Y \geq y, S = s)$$

$$u > s \quad \Rightarrow \quad 0 \leq P[Y(s) = y|Y(s) \geq y, S = u] \leq P(Y = y|Y \geq y, S = s)$$

and

$$\begin{aligned} u < s \quad &\Rightarrow \quad 0 \leq P[Y(s) = y'|Y(s) \geq y', S = u] \leq P(Y = y'|Y \geq y', S = s) \qquad (19) \\ u = s \quad &\Rightarrow \quad P[Y(s) = y'|Y(s) \geq y', S = u] = P(Y = y'|Y \geq y', S = s) \\ u > s \quad &\Rightarrow \quad P(Y = y'|Y \geq y', S = s) \leq P[Y(s) = y'|Y(s) \geq y', S = u] \leq 1 \end{aligned}$$

In the absence of further information, these bounds are sharp.

Corollary 1 provides informative upper or lower bounds on the counterfactual transition probabilities, and point identification if $u = s$. The bounds are informative also with respect to the unconditional distribution $P[Y(s)]$, $\forall s \in \mathcal{S}$ following the arguments outlined in Section 2. The MIV and PMIV assumptions with $V = S$ thus have identifying power on the distributions of interest, but relax the standard exogeneity assumptions often imposed in the literature. The following section motivates the PMIV assumptions within a structural model context and illustrates the construction of bounds in a simple simulated data environment.

4 Illustrations

4.1 Simulated Data Example

Suppose a simple structural model with two treatment states ($\mathcal{S} = \{0, 1\}$), $J = 3$ outcomes, and no covariates, that reflects the information structure of agents in the decision process:

$$S = \mathcal{I}(\nu \geq \alpha) \qquad (20)$$

$$\begin{aligned} Y(s) &= 1 \quad \text{if} \quad \varepsilon_1 \leq \tau_{s,1}, \qquad s = 0, 1 \\ Y(s) &= 2 \quad \text{if} \quad \varepsilon_2 \leq \tau_{s,2} \quad \text{and} \quad \varepsilon_1 > \tau_{s,1} \\ Y(s) &= 3 \quad \text{if} \quad \varepsilon_2 > \tau_{s,2} \quad \text{and} \quad \varepsilon_1 > \tau_{s,1}, \end{aligned}$$

where $\mathcal{I}(A)$ is the logical indicator function that returns 1 if A is true. The selection process is modelled as a binary choice problem, the outcome process is modelled as a sequence of binary choices, one for each step, conditional on the continuation up to this step. Only the pair (Y,S) is observed, where $Y = Y(S)$.

The structural model in (20) does not impose *a-priori* restrictions on the joint distribution of $(\nu, \varepsilon_1, \varepsilon_2)$, and in particular on the correlations between the error term of the selection process and the error terms of the outcome processes (except for common regularity conditions such as finite and positive definite covariance matrix). Moreover, the model allows for arbitrary behavioral assumptions on the sequential decisions since no *a-priori* assumptions are imposed on the correlation between the error terms in each step. The parameters α and τ are thresholds that determine the propensity to be selected into one of the two states, and the likelihood of dropping out of the sequential process in a perticular stage in each of the two states.

The model as presented in (20) imposes the restriction that the error terms in both states are generated by the same process. This assumption substantially simplifies the illustration below, by allowing to focus on the influence of the main parameters on the construction and properties of the bounds, but it can be relaxed in further analyses such that ε does also carry a subscript s. The model can be extended to allow for covariates X by making α and $\tau_{s,y}$, $\forall s, y$ functions of X. However, in line with the notion of the paper, the model avoids exclusion restrictions (i.e, non-overlapping subsets of X appear in α and τ), such that accounting for covariates amounts to defining treatment effects (and deriving bounds) for subpopulations described by X.

In order to illustrate the construction of bounds, the error terms $(\nu, \varepsilon_1, \varepsilon_2)$ are assumed jointly normal with zero means and covariance matrix

$$\Sigma = \begin{pmatrix} 1 & \sigma_{\nu,1} & \sigma_{\nu,2} \\ \sigma_{\nu,1} & 1 & \sigma_{1,2} \\ \sigma_{\nu,2} & \sigma_{1,2} & 1 \end{pmatrix}.$$

The threshold values in state $s = 0$ are fixed at $\tau_{0,1} = 0.8$ and $\tau_{0,2} = 0.3$, and the construction of bounds is illustrated if one of α, $\sigma_{\nu,1}$, $\sigma_{\nu,2}$, $\sigma_{1,2}$, $\tau_{1,1}$, or $\tau_{1,2}$ is varied. Table 1 provides an overview of the design. Each of the following six figures shows the true potential outcome distribution, the no-assumptions lower and upper bounds, and the lower and upper bounds of

Proposition 2/Corollary 1, assuming no uncertainty about the sign of the inequalities.

— Insert Table 1 about here —

Figure 1 displays the potential outcome distribution if α is varied from -2 to 2 (and all other parameters are fixed as shown in Table 1). Here, and in the following figures, the upper two diagrams show the probability that the potential outcome takes value 1 in state 1 (on the left) and in state 0 (on the right). The middle and bottom diagrams show the according probabilities for outcomes 2 and 3, respectively. Since α does only affect the selection status but does not affect the potential outcome process, the true outcome probabilities in each of the two states are constant (indicated by the thick black line). The construction and the properties of the bounds, however, depend on the value of α because only (Y, S) is observed.

— Insert Figure 1 about here —

Consider outcome 1 in state 1. The potential outcome probability can be written as $P[Y(1) = 1] = P(S = 1, Y = 1) + P[Y(1) = 1|S = 0]P(S = 0)$ by the law of total probability and the observation rule for Y. The no-assumptions bounds, indicated by the small grey circles and triangles, impose $P[Y(1) = 1|S = 0] \in (0, 1)$ by the definition of a probability. The bounds on $P[Y(1) = 1]$ are relatively tight if $P(S = 0)$ is small, and relatively wide if $P(S = 0)$ is large. The former case is associated with α small, the latter with α large, as shown in the figure. Analogous arguments hold for the no-assumptions bounds and outcome levels 2 and 3. The role of α is reversed if the potential outcome probabilities in state 0 are considered, with the bounds being relatively wide if $P(S = 1)$ is large (or α small) and the bounds being relatively tight if $P(S = 0)$ is small (or α large).

Now consider the bounds derived in Proposition 2/Corollary 1 (indicated by the large hollow circles and triangles). These bounds are often tighter, and never wider, than the no-assumptions bounds, which can be explained as follows. By the assumption of normality, $\sigma_{\nu,1} = 0.3$, and $\sigma_{\nu,2} = 0.5$ it follows that $P[Y(s) = 1|Y(s) \geq 1, S = 0] > P[Y(s) = 1|Y(s) \geq 1, S = 1]$ and $P[Y(s) = 2|Y(s) \geq 2, S = 0] > P[Y(s) = 2|Y(s) \geq 2, S = 1]$ for $s = 0, 1$.

The former is implied by $\sigma_{\nu,1}$ being positive, the latter by the joint set of assumptions. It follows from the former that $A1 \equiv P(Y = 1|Y \geq 1, S = 1)$ can be used as a lower bound for $P[Y(1) = 1|Y(1) \geq 1, S = 0]$ instead of zero, and $B1 \equiv P(Y = 1|Y \geq 1, S = 0)$ can be used as an upper bound for $P[Y(0) = 1|Y(0) \geq 1, S = 1]$ instead of one. Moreover, $A2 \equiv P(Y = 2|Y \geq 2, S = 1)$ can be used as a lower bound for $P[Y(1) = 2|Y(1) \geq 2, S = 0]$ instead of zero, and $B2 \equiv P(Y = 2|Y \geq 2, S = 0)$ can be used as an upper bound for $P[Y(0) = 2|Y(0) \geq 2, S = 1]$ instead of one.

Since the counterfactual probability for outcome 1 in state 1 can be written as $P[Y(1) = 1|S = 0] = P[Y(1) = 1|Y(1) \geq 1, S = 0]$, it lies in the interval $(A1, 1)$ by the argument above. Thus, the lower bound on $P[Y(1) = 1]$ is shifted upwards compared to the no-assumptions lower bound, and the upper bound remains the same. For outcome 2, write $P[Y(1) = 2|S = 0] = P[Y(1) = 2|Y(1) \geq 2, S = 0]\{1 - P[Y(1) = 1|Y(1) \geq 1, S = 0]\}$ which lies in the interval $(0, A2(1 - A1))$. Thus, the upper bound on $P[Y(1) = 2]$ is smaller than the no-assumptions upper bound and the lower bound remains the same. The potential outcome probability for the highest category 3 can be written as $P[Y(1) = 3|S = 0] = \{1 - P[Y(1) = 2|Y(1) \geq 2, S = 0]\}\{1 - P[Y(1) = 1|Y(1) \geq 1, S = 0]\}$ which lies in the interval $(0, (1 - A1)(1 - A2))$ by the assumptions above, and thus again only the upper bound is shifted downwards.

Analogous arguments hold for the three outcomes in state 0. The counterfactual probability for outcome 1 lies in the interval $(0, B1)$, it lies in the interval $(0, B2)$ for outcome 2, and in the interval $((1 - B1)(1 - B2), 1)$ for outcome 3. Thus, in comparison to the no-assumptions bounds, the upper bounds as implied by Proposition 2/Corollary 1 are shifted downwards for the outcome levels 1 and 2, and the lower bound is shifted upwards for the outcome level 3. The magnitude of the shift depends on the magnitude of $B1$ and $B2$, as well as the selection probability $P(S = 1)$ (which depends on α).

In the setup of the simulation, the width of the bounds as implied by Proposition 2/Corollary 1 is substantially smaller than the no-assumptions bounds for all outcomes in state 1, while the information gain in terms of width is less pronounced for state 0. This result is not determined by a single parameter but by the overall setup of the study and the imposed assumptions on the parameters.

Figure 2 displays the results if the correlation between the error term in the selection process and the error term of the first step in the sequential process, $\sigma_{\nu,1}$, is varied from -0.5 to 0.5. If there is no selection on unobservables in the first step of the sequential process, i.e., if $\sigma_{\nu,1} = 0$, then the counterfactual probabilities become point identified because $P[Y(s) = 1|S = 0] = P[Y(s) = 1|S = 1]$ for $s = 0, 1$. This is indicated in the graph at that point when the large hollow circle and triangle coincide with the true outcome probability (the thick black line). If $\sigma_{\nu,1}$ is negative (positive), then

$$P[Y(s) = 1|Y(s) \geq 1, S = 0] < (>) P[Y(s) = 1|Y(s) \geq 1, S = 1], \quad s = 0, 1$$

In all cases, $P[Y(s) = 2|Y(s) \geq 2, S = 0] > P[Y(s) = 2|Y(s) \geq 2, S = 1]$ for $s = 0, 1$ by the assumptions of the setup. This example is interesting for the comparison of PMTS and MTS. If $\sigma_{\nu,1}$ is positive, then the two are equivalent. However, if $\sigma_{\nu,1}$ is negative, then the MTS assumption is violated while the PMTS assumption is not.

— Insert Figure 2 about here —

According to Proposition 2/Corollary 1, the counterfactual probabilities in state 1 and in the case of $\sigma_{\nu,1}$ negative can be bounded by $P[Y(1) = 1|S = 0] \in (0, A1)$, $P[Y(1) = 2|S = 0] \in (A2(1 - A1), 1)$, and $P[Y(1) = 3|S = 0] \in (0, 1 - A2)$, and in state 0, the bounds can be derived as $P[Y(0) = 1|S = 1] \in (B1, 1)$, $P[Y(1) = 2|S = 0] \in (0, B2(1 - B1))$, and $P[Y(1) = 3|S = 0] \in (0, 1 - B1)$. If $\sigma_{\nu,1}$ is positive, then the bounds are the same as the ones described above when α is varied.

Comparing the two cases $\sigma_{\nu,1}$ positive and $\sigma_{\nu,1}$ negative, it can be observed that there is no uni-directional shift in the bounds compared to the no-assumptions case. The sign and magnitude of the shift depends on all parameters such that no general conclusion can be drawn, even from this simple model. However, compared to the no-assumptions case, the information gain can be substantial, cutting the possible range of the potential outcome distribution by more than a third (e.g., for outcomes 2 and 3 in state 0). Thus, while the weak monotonicity assumption may not always have point-identifying power, it is nevertheless informative regarding the outcome process. As a further result, the bounds derived by wrongly imposing the MTS assumption may not always include the true probability given PMTS. In the given example, this happens for $P[Y(1)=3]$ and $P[Y(0)=3]$, and $\sigma_{\nu,1}$ about -0.2.

Figure 3 displays the results when $\sigma_{\nu,2}$ is varied from -0.5 to 0.5. A variation in $\sigma_{\nu,2}$ does not alter the first stage in the sequential process, and since $\sigma_{\nu,1}$ is assumed to be positive (Table 1), the potential transition probabilities fulfill the assumption $P[Y(s)=1|Y(s)\geq 1, S=0] > P[Y(s)=1|Y(s)\geq 1, S=1]$, $s=0,1$. The setup is chosen such that there is a unique shift in the sign of the inequalities for the potential transition probabilities for the second category at $\sigma_{\nu,2}\approx 0.112$. If $\sigma_{\nu,2}<0.112$, then $P[Y(s)=1|Y(s)\geq 1, S=0] < P[Y(s)=1|Y(s)\geq 1, S=1]$ for $s=0,1$, and if $\sigma_{\nu,2}>0.112$, then $P[Y(s)=1|Y(s)\geq 1, S=0] > P[Y(s)=1|Y(s)\geq 1, S=1]$ for $s=0,1$. While the latter case is analogous to the scenario when α is varied, the former case implies the following bounds on the counterfactual probabilities:

$$
\begin{aligned}
P[Y(1)=1|S=0] &\in (A1, 1)\\
P[Y(1)=2|S=0] &\in (0, A2(1-A1))\\
P[Y(1)=3|S=0] &\in (0, 1-A1)
\end{aligned}
$$

and

$$
\begin{aligned}
P[Y(0)=1|S=1] &\in (0, B1)\\
P[Y(0)=2|S=1] &\in (B2(1-B1), 1)\\
P[Y(0)=3|S=1] &\in (0, 1-B2)
\end{aligned}
$$

Thus, in state 1 the lower bound is shifted upwards for outcome 1, and the upper bounds are shifted downwards for outcomes 2 and 3 compared to the no-assumptions bounds. In state 0, the lower bound is identical to the no-assumptions lower bound for outcomes 1 and 3 while the upper bound is smaller, and the opposite holds for outcome 2.

— Insert Figure 3 about here —

The different role of the information set in the construction of bounds should be noted when comparing Figures 2 and 3. For example, the sign of the correlation between the first step error term and the error term of the selection process affects the bounds in all outcome categories. If economic theory is informative regarding the sign of the correlation, then this affects the information available on the entire outcome distribution. The correlation between the second step error term and the selection error term does not affect the probability of the lower outcome 1 (due to the sequential structure of the model), and thus is only informative on the same and the subsequent categories. The magnitude of the information gain depends on the particular characteristics of the model (i.e., the covariance matrix, the threshold parameters, and the number of outcomes), but any weak information that can be credibly imposed on these determinants can be successfully employed to derive bounds on the potential outcome distribution that are tighter than the no-assumptions bounds.

The last three figures display the results when the correlation between the first step and the second step error term ($\sigma_{1,2}$) is varied (Figure 4), and the threshold parameters $\tau_{1,1}$ and $\tau_{1,2}$ are varied (Figures 5 and 6). In all cases, the setup of the study ensures conformability with the PMIV assumptions in the sense that $P[Y(s) = y|Y(s) \geq y, S = 0] > P[Y(s) = y|Y(s) \geq y, S = 1]$ for $s = 0, 1$ and $y = 1, 2$ such that the construction of bounds is analogous to the case when α is varied. It should be noted that while $\tau_{1,1}$ and $\tau_{1,2}$ are the threshold parameters in state 1, they also affect the bounds on the potential outcome distribution in state 0 because they affect what is observed by the researcher and what assumptions can be imposed on the potential conditional transition probabilities. While this information does not vary with $\tau_{1,1}$ and $\tau_{1,2}$, it is already sufficient to obtain bounds that are more informative than the no-assumptions bounds.

— Insert Figures 4 to 6 about here —

If X covariates are present, then the construction of the bounds as presented above is local (conditional on X) and the overall bounds (unconditional on X) can be obtained as a weighted average. The PMIV assumptions that can be imposed in each sub-population described by X may differ such that either more informative upper bounds or more informative lower bounds can be imposed, conditional on X. It is also consistent with the concept of the PMIV assumptions if for some sub-populations no additional information can be invoked and hence the no-assumptions bounds still apply.

4.2 Schooling and Fertility

As an empirical example, I will consider how the number of children ever born to a woman varies with her level of schooling. I will impose weak assumptions on the data-generating process that are consistent with the related literature. These assumptions are also consistent with the PMIV assumptions, as discussed below. However, compared to most previous analyses, I will neither make distributional or functional form assumptions, nor invoke strong independence or exclusion restrictions. Instead, I will discuss the implications of the PMIV assumptions on the counterfactual distributions of interest.

Unfortunately, economic theory alone does not provide a clear-cut prediction on the sign of the correlation between education and fertility. In the traditional perspective, more educated women face higher opportunity costs of childbearing, which on average reduces the expected number of children, but they also tend to have higher wages, and/or a higher educated partner with a higher income (Becker 1991, Pollak and Watkins 1993), which may contribute to a higher fertility of the more educated women (although the latter conclusion is not without debate, see for example Becker and Lewis 1973 for a discussion on the quality-quantity aspect). In addition, higher educated women tend to postpone their marriage and fertility decisions, they tend to have different views about family formation and female autonomy, and also tend to have a better knowledge about fertility control (Boulier and Rosenzweig 1984, Rosenzweig and Schultz 1985, Mason 1986, Schultz 1993, Cheng and Nwachukwu 1997, Lam and Duryea 1999, Basu 2002, Bratti 2003), which are all factors that may reduce the number of children of the more educated women.

The empirical findings on the correlation between education and fertility mainly point to a negative sign (Rosenzweig and Schultz 1985, Easterlin 1987, Becker 1991, Sander 1992, Schultz 1993, Lam and Duryea 1999). However, some studies provide evidence in favor of a positive association (e.g., Moffitt 1984), although the effects are weakly significant or insignificant.

In the following analysis, I will combine the results from the above cited literature (mainly the model of Boulier and Rosenzweig 1984) with the literature on fertility expectations and ideal family size (e.g., Blake 1974, Coombs 1978, Ajzen (1985), and Schoen *et al.* 1999). In this view, fertility decisions follow a sequential process and the effect of schooling depends on the actual number of children. The higher educated women more often stay childless than the lower educated women, on average. However, given already one child, i.e., the decision in favor of getting children, the higher educated women will on average be better able to achieve ideal family size goals. In the US, where the employed data come from, this amounts to an ideal of two children (the median and modal value over time). Thus, the decision for one child as opposed to two or more children is more likely for the low educated women than for the high educated women, but the conditional transition probabilities for the second and more children will be higher for the more than the lower educated women. Formally,

$$\begin{aligned} P(Y(s) = 0|S = u_1) &\leq P(Y(s) = 0|S = u_2) \\ P(Y(s) = 1|Y(s) \geq 1, S = u_1) &\geq P(Y(s) = 1|Y(s) \geq 1, S = u_2) \\ P(Y(s) = y|Y(s) \geq y, S = u_1) &\leq P(Y(s) = y|Y(s) \geq y, S = u_2), \quad y \geq 2 \end{aligned}$$

where $Y(s)$ is the number of children if the woman were to receive s years of schooling, S is the realized schooling, and $u_1 < u_2$.

The data used for the analysis are taken from the General Social Survey (GSS) 1972-2008 (see the website of the National Opinion Research Center www.norc.org for further details on the data). I restrict attention to the (sub-) population of white women aged 55-70 that are not in the labor force. Table 2 gives the estimates of the schooling distribution in a sample of 2,976 randomly selected women from this population, with the schooling variable aggregated to four categories: less than 12 years, 12 years, 13-15 years, and 16 or more years of schooling.

— Insert Table 2 about here —

The estimates of the conditional transition probabilities in the fertility distribution, for given levels of schooling, are also shown in Table 2 (the numbers of observations used to estimate each quantity are provided in square brackets). The observed drop-out rates (fertility transitions) tend to increase with the education level for each but the first child. This may be interpreted as an indicator for the imposed PMTS assumptions, though it cannot be used as a test because the estimated numbers are based on the observed instead of the potential outcomes. Stronger assumptions on the relationship between schooling and the number of children would be necessary in order to conduct a valid test of the (combined) assumptions.

Figure 7 shows the estimated potential outcome distributions for each of the four schooling categories. The horizontal axes display the number of children y, where 5 and more children are summarized into a single category, and the vertical axes display the according probabilities $P[Y(s) = y]$. Using the empirical evidence alone the potential outcome probabilities can only be bounded with the estimated range being marked by the light grey bars.

— Insert Figure 7 about here —

The bounds under the PMTS assumptions are indicated by the black bars. For comparison, the bounds obtained under the MTS assumption (with weak inequality for the transition from one child reversed) are indicated by the grey bars. Imposing these assumptions does substantially narrow the no-assumptions bounds, although point identification of any of the potential outcome probabilities is not achieved. The shrinkage in the estimated range depends on the schooling and the outcome level. The lower PMTS bound for outcomes greater than or equal to two is always zero since, by assumption, the lower bound on at least one of the conditional transitions is zero in these cases. The lowering of the upper PMTS bound is the greater, the more the lower bound on the conditional transitions in the previous steps is raised.

Nonparametric bounds on the treatment effects, if defined as the differences in the outcome distributions, can be derived from these graphs by taking differences between the according bars. If the only information imposed on the data-generating process is the PMTS structure, then the bounds in Figure 7 are sharp (noting the additional variation that arises due to sampling error). Tighter bounds on the counterfactual distributions can only be obtained by imposing stronger assumptions.

5 Conclusion

Manski and Pepper (2000) discuss monotone instrumental variables in order to relax common IV assumptions. In a nutshell, the equality of mean responses is replaced by a weak inequality conditional on the variation in the instrument (under suitable order conditions). While the MIV assumptions generally do not point-identify the potential outcome distribution, they can be successfully explored to derive informative bounds. Manski and Pepper point to the extension of their MIV idea to stochastic dominance. To the best of my knowledge, this paper is the first that investigates the MIV concept in the context of a sequentially ordered outcome process, thereby exploring a particular form of stochastic dominance.

Assuming that agents make decisions under a sequential update of information, the paper applies the MIV idea in order to impose weak stochastic dominance on the conditional transition probabilities. Reflecting the sequential order, the sign of stochastic dominance might depend on the outcome level itself. This concept is well-founded in economic demand models with heterogenous preferences and where income and substitution effects interact. Since the paper does not impose independence assumptions or exclusion restrictions, the results derived above have implications for a large number of applications where no such assumptions can be credibly imposed, but where a sequential process can be motivated and weak information about the direction of effects is available.

Further research might combine some sort of (weak) independence condition with the structural model imposed in Section 4 in order to gain further insights on the sign of the inequalities in the PMIV assumptions. Moreover, two issues should receive attention regarding inference.

First, the potential finite sample bias in the analogue estimates needs to be appropriately corrected (see also Kreider and Pepper 2007, Manski and Pepper 2009). Second, the uncertainty in the sign of the weak inequalities in the PMIV assumptions must be accounted for in order to derive confidence intervals with a pre-defined coverage probability.

References

Abbring, J.H. and J.J. Heckman (2007), Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation, in: J.J. Heckman and E. Leamer (eds.) *Handbook of Econometrics*, Volume 6B, Elsevier, 5145-5303.

Abbring, J.H. and G.J. Van den Berg (2003), The nonparametric identification of treatment effects in duration models, *Econometrica*, 71, 1491-1517.

Ajzen, I. (1985), From intentions to actions: A theory of planned behavior, in: J. Kuhl and J. Beckmann (eds.) *Action control: From cognition to behavior*, Springer, 11-39.

Amemiya, T. (1975), Qualitative response models, *Annals of Economic and Social Measurement*, 4, 363-372.

Amemiya, T. (1985), *Advanced Econometrics*, Harvard University Press.

Angrist, J.D., G.W. Imbens and D. Rubin (1996), Identification of causal effects using instrumental variables, *Journal of the American Statistical Association*, 91, 444-455.

Basu, A.M. (2002), Why Does Education Lead to Lower Fertility? A Critical Review of Some of the Possibilities, *World Development*, 30, 1779-1790.

Becker, G.S. (1991), *A Treatise on the Family*, Harvard University Press.

Becker, G.S. and H.G. Lewis (1973), On the Interaction between the Quantity and Quality of Children, *Journal of Political Economy*, 81, S279-288.

Blake, J. (1974), Can We Believe Recent Data on Birth Expectations in the United States? *Demography*, 11, 25-44.

Boes, S. (2007), Nonparametric Analysis of Treatment Effects in Ordered Response Models, *SOI Working Paper No. 0709.*

Boes, S. (2009), Bounds on Ordered Counterfactual Distributions Under Weak Monotonicity Constraints, *unpublished manuscript.*

Boulier, B.L. and M.R. Rosenzweig (1984), Schooling, Search, and Spouse Selection: Testing Economic Theories of Marriage and Household Behavior, *The Journal of Political Economy*, 92, 712-732.

Bratti, M. (2003), Labour Force Participation and Marital Fertility of Italian Women: The Role of Education, *Journal of Population Economics*, 16, 525-554.

Brooks-Gunn, J. (2003). Do you believe in magic? What we can expect from early childhood intervention programs. *Social Policy Report, Society for Research in Child Development*, 17(1). Available online: http://www.srcd.org/Documents/ Publications/SPR/spr17-1.pdf.

Carneiro, P., K. Hansen and J.J. Heckman (2003), Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice, *International Economic Review*, 44, 361-422.

Coombs, L.C. (1978), How Many Children Do Couples Really Want? *Family Planning Perspectives*, 10, 303-308.

Cheng, B.S. and S.L.S. Nwachukwu (1997), The Effect of Education on Fertility in Taiwan: A Time Series Analysis, *Economics Letters*, 56, 95-99.

Cunha, F., J.J. Heckman and S. Navarro (2007), The identification and economic content of ordered choice models with stochastic thresholds, *International Economic Review*, 48, 1273-1309.

Easterlin, R.A. (1987), Fertility, in J. Eatwell, M. Milgate and P. Newman (eds.) *The New Palgrave: A Dictionary of Economics 2*, The Stockton Press, 302-308.

Gill, R.D. and J.M. Robins (2001), Causal inference for complex longitudinal data: The continuous case, *The Annals of Statistics*, 29, 1785-1811.

Haavelmo, T. (1943), The statistical implications of a system of simultaneous equations, *Econometrica*, 11, 1-12.

Heckman, J.J. (2001), Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture, *Journal of Political Economy*, 109, 673-748.

Heckman, J.J. and B.E. Honoré (1990), The empirical content of the Roy model, *Econometrica*, 58, 1121-1149.

Heckman, J.J. and S. Navarro (2007), Dynamic Discrete Choice and Dynamic Treatment Effects, *Journal of Econometrics*, 136, 341-396.

Heckman, J.J and R.J. Willis (1977), A Beta-logistic Model for the Analysis of Sequential Labor Force Participation by Married Women, *Journal of Political Economy*, 85, 27-58.

Heckman, J.J. and E.J. Vytlacil (2005), Structural equations, treatment effects and econometric policy evaluation, *Econometrica*, 73, 669-738.

Heckman, J.J. and E.J. Vytlacil (2007a), Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation, in: J.J. Heckman and E. Leamer (eds.) *Handbook of Econometrics*, Volume 6B, Elsevier, 4779-4874.

Heckman, J.J. and E.J. Vytlacil (2007b), Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative economic estimators to evaluate social programs and to forecast their effects in new environments, in: J.J. Heckman and E. Leamer (eds.) *Handbook of Econometrics*, Volume 6B, Elsevier, 4875-5144.

Imbens, G.W. and J.D. Angrist (1994), Identification and Estimation of Local Average Treatment Effects, *Econometrica*, 62, 467-476.

Kahn, L.M. and K. Morimune (1979), Unions and employment stability: a sequential logit approach, *International Economic Review*, 20, 217-235.

Kreider, B. and J. Pepper (2007), Disability and employment: Reevaluating the evidence in light of reporting errors, *Journal of the American Statistical Association*, 102, 432441.

Lam, D. and S. Duryea (1999), Effects of Schooling on Fertility, Labor Supply and Investments in Children, with Evidence from Brazil. *Journal of Human Resources*, 34, 160-192.

Lancaster, T. (1992), *The Econometric Analysis of Transition Data*, Econometrics Society Monographs, Cambridge University Press.

Lechner, M. (2008), Matching estimation of dynamic treatment models: Some practical issues, in: D. Millimet, J. Smith, and E. Vytlacil (eds.), *Advances in Econometrics*, Volume 21, Modelling and Evaluating Treatment Effects in Econometrics, 289-333.

Lechner, M. (2009), Sequential Causal Models for the Evaluation of Labor Market Programs, *Journal of Business & Economic Statistics*, 27, 71-83.

Lechner, M. and R. Miquel (2009), Identification of the Effects of Dynamic Treatments by Sequential Conditional Independence Assumptions, *Empirical Economics*, forthcoming.

Manski, C.F. (1990), Nonparametric Bounds on Treatment Effects, *American Economic Review, Papers and Proceedings*, 80, 319-323.

Manski, C.F. (1994), The Selection Problem, in: C. Manski and D. McFadden (eds.) *Advances in Econometrics: Sixth World Congress*, Cambridge University Press, Cambridge.

Manski, C.F. (1997), Monotone Treatment Response, *Econometrica*, 65, 1311-1334.

Manski, C.F. (2000), Identification problems and decisions under ambiguity: empirical analysis of treatment response and normative analysis of treatment choice, *Journal of Econometrics*, 95, 415-442.

Manski, C.F. (2007), *Identification for Prediction and Decision*, Harvard University Press.

Manski, C.F. and J. Pepper (2000), Monotone instrumental variables: with an application to the returns to schooling, *Econometrica*, 68, 997-1010.

Manski, C.F. and J. Pepper (2009), More on monotone instrumental variables, *Econometrics Journal*, S1, S200-S216.

Mason, K.O. (1986), The Status of Women: Conceptual and Methodological Issues in Demographic Studies. *Sociological Forum*, 1, 284-300.

Pollak, R.A. and S.C. Watkins (1993), Cultural and Economic Approaches to Fertility: Proper Marriage or Mesalliance? *Population and Development Review*, 19, 467-496.

Robins, J.M. (1989), The analysis of randomized and non-randomized aids treatment trials using a new approach to causal inference in longitudinal studies, in: L. Sechrest, H.

Freeman, A. Mulley (eds.), *Health Services Research Methodology: A Focus on AIDS*. U.S. Department of Health and Human Services, National Center for Health Services Research and Health Care Technology Assessment, Rockville, MD, 113-159.

Robins, J.M. (1997), Causal inference from complex longitudinal data, in: M. Berkane (ed.), *Latent Variable Modeling and Applications to Causality.* Lecture Notes in Statistics. Springer-Verlag, New York, 69-117.

Rosenzweig, M.R. and T.P. Schultz (1985), The Demand for and Supply of Births: Fertility and Its Life Cycle Consequences. *American Economic Review*, 75, 992-1015.

Roy, A. (1951), Some thoughts on the distribution of earnings, *Oxford Economic Papers*, 3, 135-146.

Rubin, D.B. (1974), Estimating causal effects of treatments in randomized and nonrandomized studies, *Journal of Educational Psychology*, 66, 688-701.

Sander, W. (1992), The effect of women's schooling on fertility, *Economics Letters*, 40, 229-33.

Schoen, R., N.M. Astone, Y.J. Kim, C.A. Nathanson and J.M. Fields (1999), Do Fertility Intentions Affect Fertility Behavior? *Journal of Marriage and the Family*, 61, 790-799.

Schultz, T.P. (1993), Returns to Women's Education, in: E.M. King and M.A. Hill (eds.) *Women's Education in Developing Countries: Barriers, Benefits, and Policies.* Johns Hopkins University Press for the World Bank, 51-99.

Shaikh, A.M. and E. Vytlacil (2005), Threshold Crossing Models and Bounds on Treatment Effects: A Nonparametric Analysis, *unpublished working paper.*

Steele, F., W. Sigle-Rushton, and O. Kravdal (2009), Consequences of Family Disruption on Childrens Educational Outcomes in Norway. *Demography*, forthcoming.

Zhang, J. (1994), Socioeconomic determinants of fertility in Hebei Province, China: an application of the sequential logit model, *Economic Development and Cultural Change*, 43, 67-90.

Figures and Tables

Table 1: Design Simulation Study

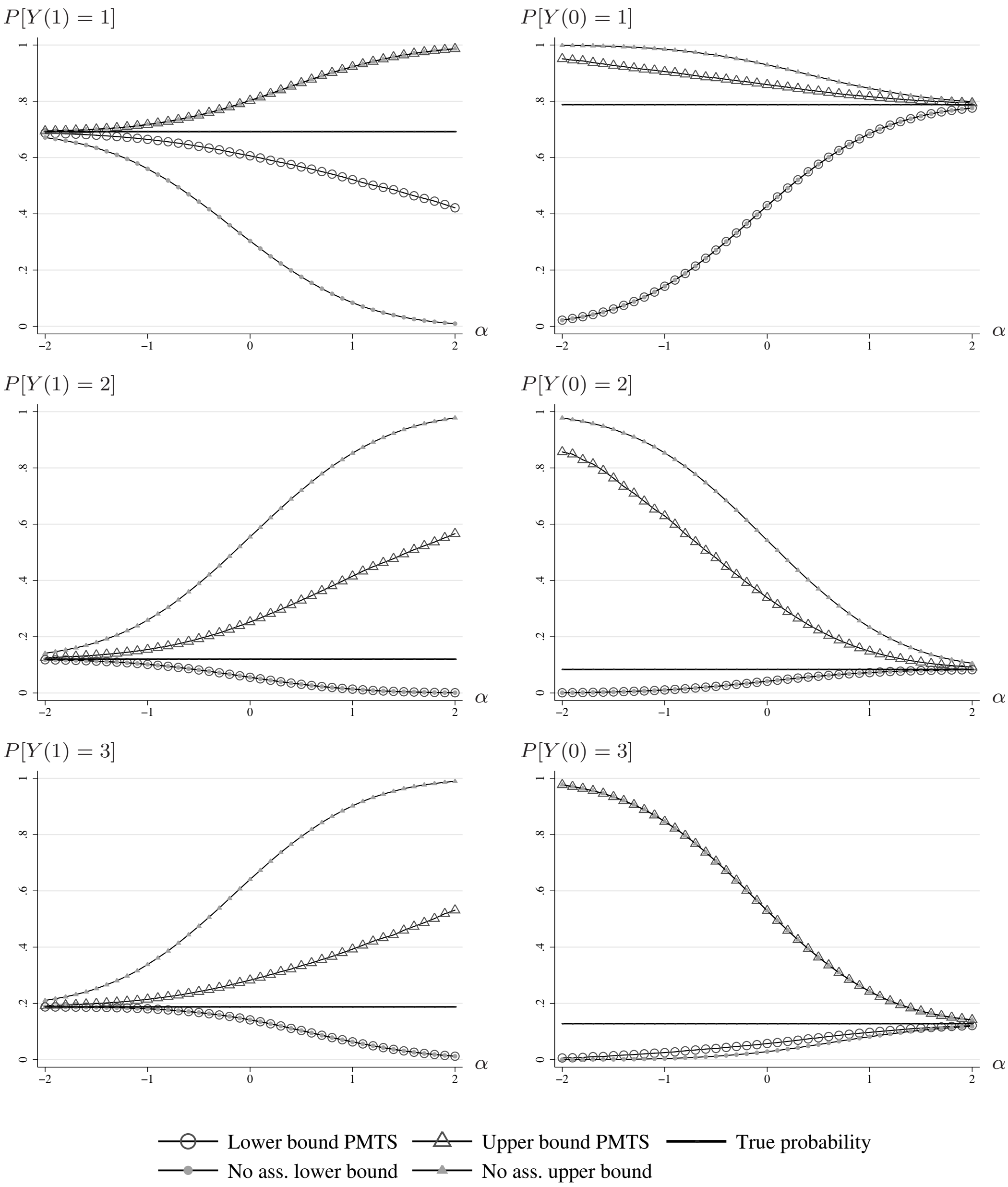
		α	$\sigma_{\nu,1}$	$\sigma_{\nu,2}$	$\sigma_{1,2}$	$\tau_{1,1}$	$\tau_{1,2}$
Figure	1	-2 to 2	0.3	0.5	0.4	0.5	0
	2	0	-0.5 to 0.5	0.5	0.4	0.5	0
	3	0	0.3	-0.5 to 0.5	0.4	0.8	0
	4	0	0.3	0.5	-0.5 to 0.5	0.5	0
	5	0	0.3	0.5	0.4	-2 to 2	0
	6	0	0.3	0.5	0.4	0.5	-2 to 2

Table 2: Schooling and Fertility - Empirical Distribution Functions

	Years of schooling (S)			
	≤ 11	12	13-15	≥ 16
$P(S)$	0.364	0.367	0.160	0.109
	[1,082]	[1,092]	[476]	[326]
Conditional transitions in the fertility distribution				
$P(Y=0\|S)$	0.106	0.127	0.149	0.221
	[1,082]	[1,092]	[476]	[326]
$P(Y=1\|Y\geq 1,S)$	0.155	0.162	0.131	0.126
	[967]	[953]	[405]	[254]
$P(Y=2\|Y\geq 2,S)$	0.285	0.362	0.367	0.441
	[817]	[799]	[352]	[222]
$P(Y=3\|Y\geq 3,S)$	0.355	0.488	0.498	0.629
	[584]	[510]	[223]	[124]
$P(Y=4\|Y\geq 4,S)$	0.377	0.529	0.580	0.522
	[377]	[261]	[112]	[46]

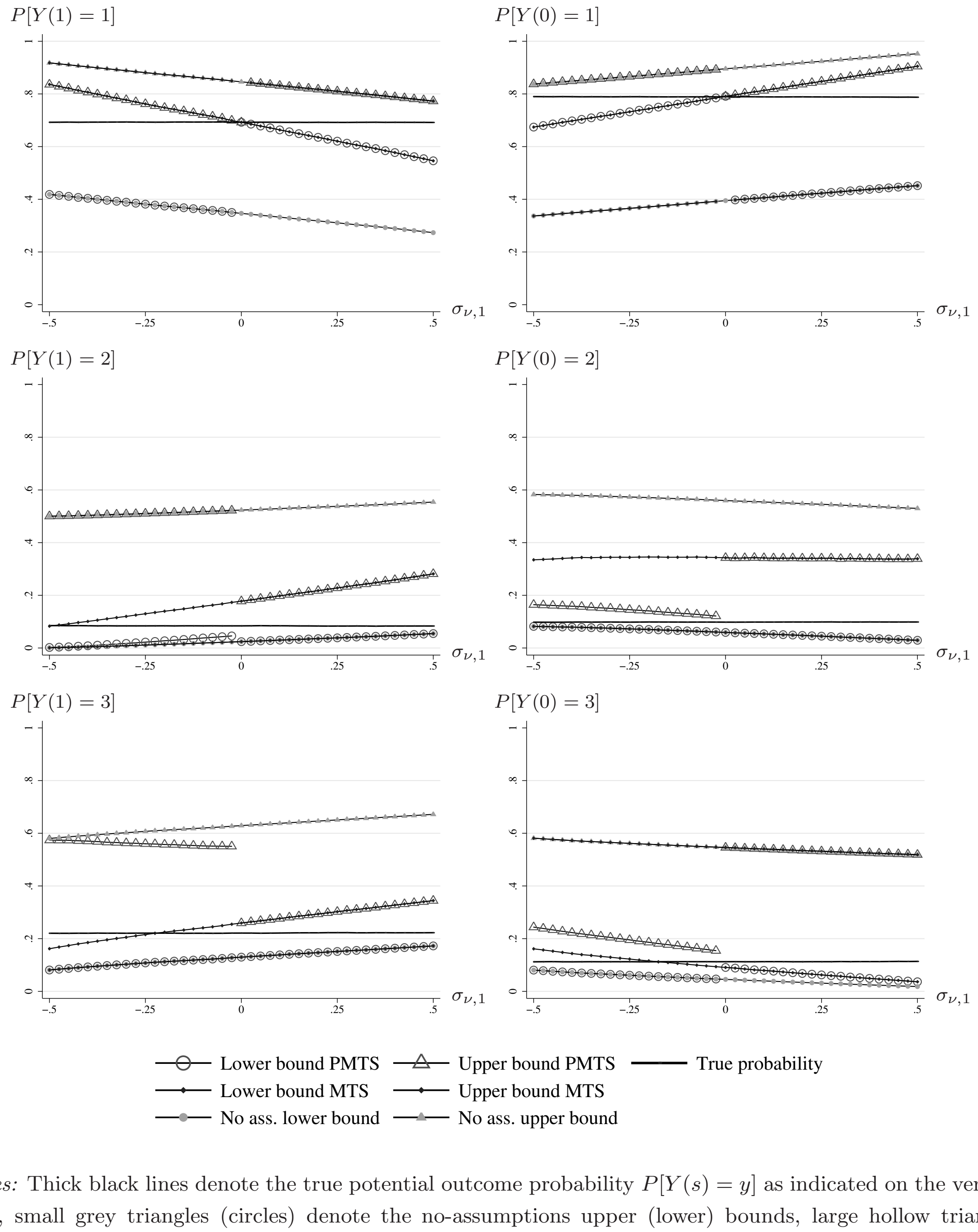
Source: GSS 1972-2008, own calculations. *Notes:* The estimates are based on a random sample of 2,976 observations on white women aged 55-70 not in the labor force. The numbers of observations used to estimate each probability are reported in square brackets. Y denotes the number of children.

Figure 1: Bounds on Potential Outcome Distributions – Variation of α



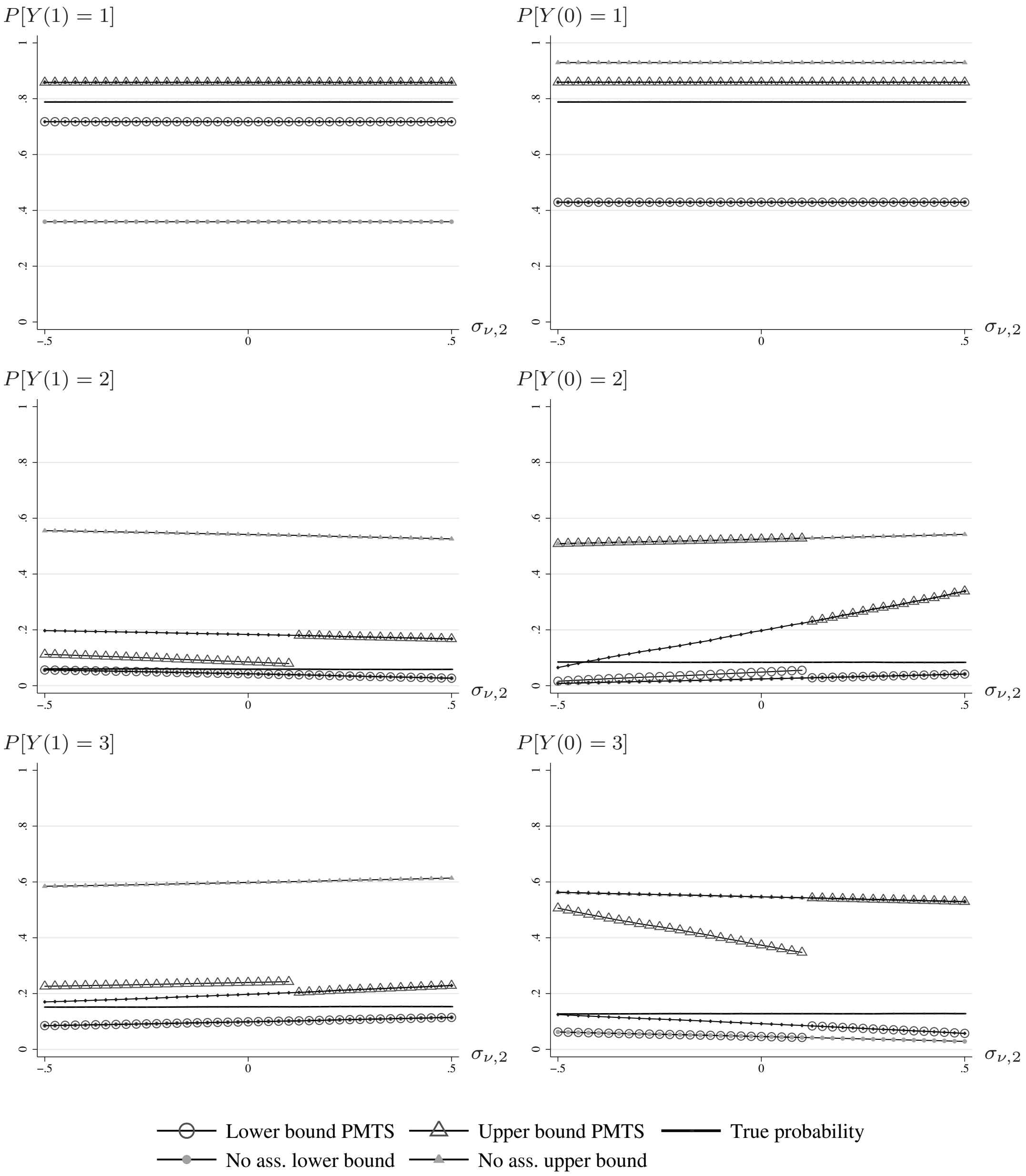
Notes: Thick black lines denote the true potential outcome probability $P[Y(s) = y]$ as indicated on the vertical axes, small grey triangles (circles) denote the no-assumptions upper (lower) bounds, large hollow triangles (circles) denote the upper (lower) bounds as derived in Proposition 1/Corollary 1 under the partial monotone treatment selection (PMTS) assumptions.

Figure 2: Bounds on Potential Outcome Distributions – Variation of $\sigma_{\nu,1}$



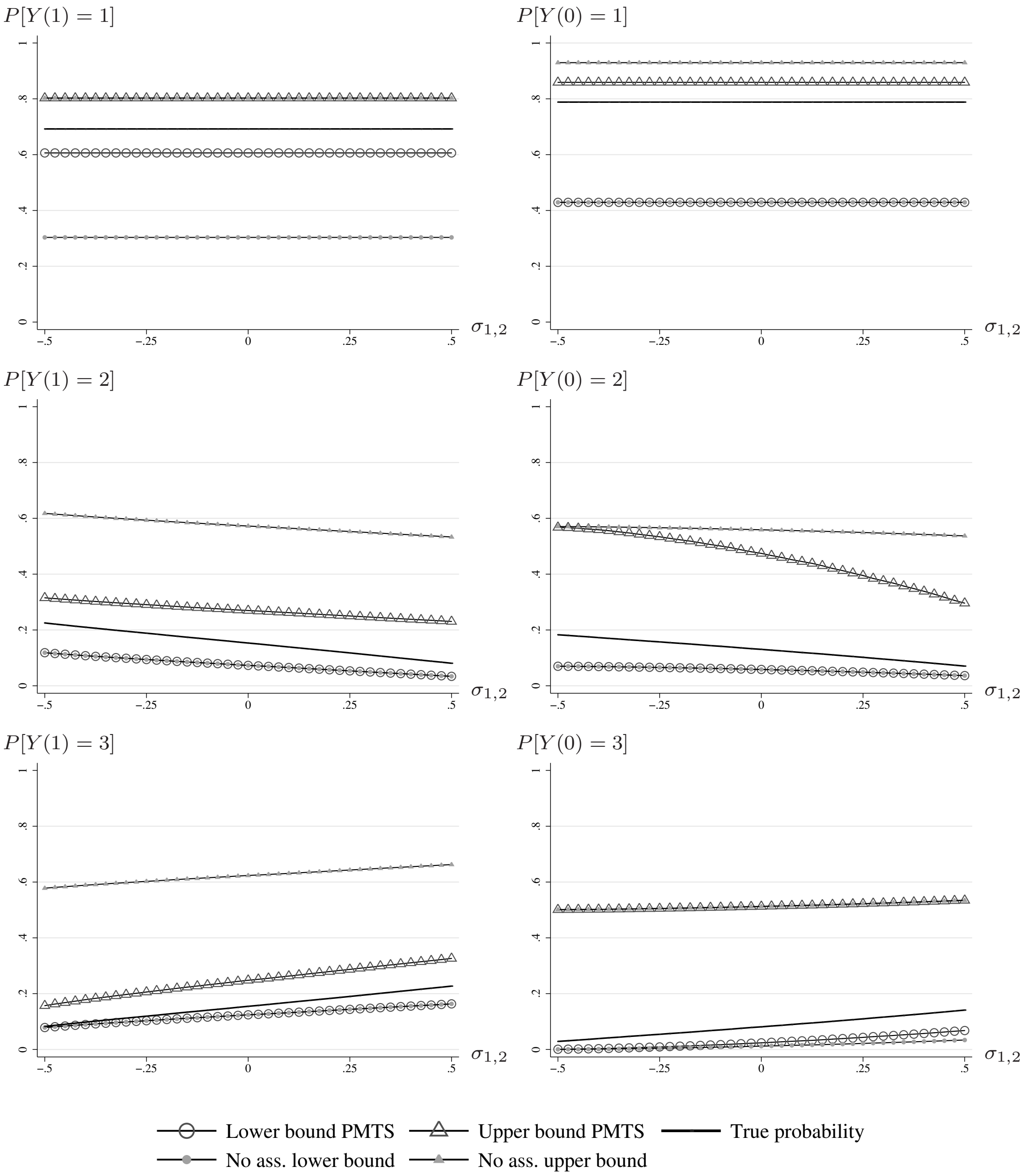
Notes: Thick black lines denote the true potential outcome probability $P[Y(s) = y]$ as indicated on the vertical axes, small grey triangles (circles) denote the no-assumptions upper (lower) bounds, large hollow triangles (circles) denote the upper (lower) bounds as derived in Proposition 1/Corollary 1 under the partial monotone treatment selection (PMTS) assumptions. For comparison, the graph also shows the bounds obtained if a monotone treatment selection (MTS) assumption is imposed (dark diamonds).

Figure 3: Bounds on Potential Outcome Distributions – Variation of $\sigma_{\nu,2}$



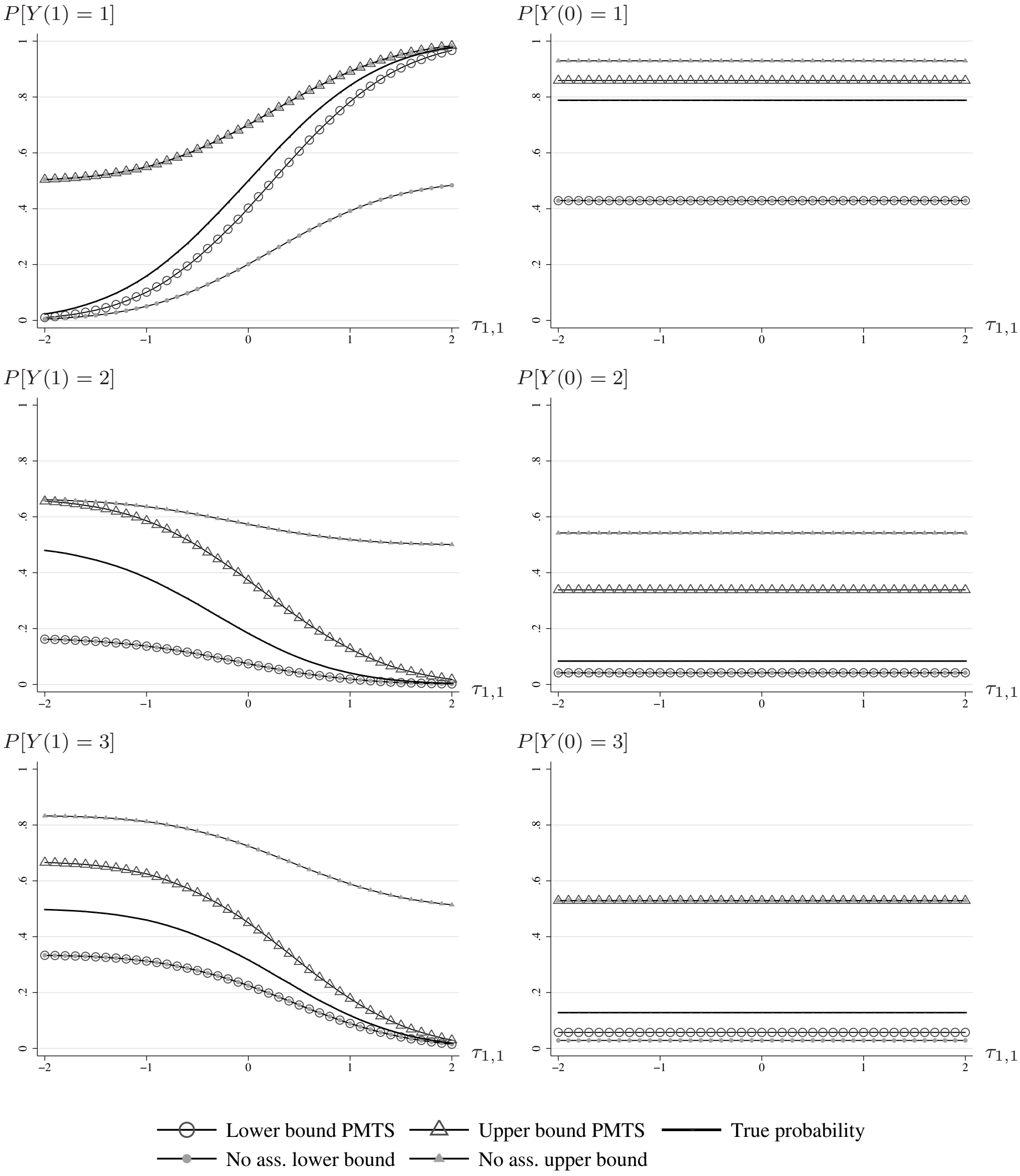
Notes: Thick black lines denote the true potential outcome probability $P[Y(s) = y]$ as indicated on the vertical axes, small grey triangles (circles) denote the no-assumptions upper (lower) bounds, large hollow triangles (circles) denote the upper (lower) bounds as derived in Proposition 1/Corollary 1 under the partial monotone treatment selection (PMTS) assumptions. For comparison, the graph also shows the bounds obtained if a monotone treatment selection (MTS) assumption is imposed (dark diamonds).

Figure 4: Bounds on Potential Outcome Distributions – Variation of $\sigma_{1,2}$



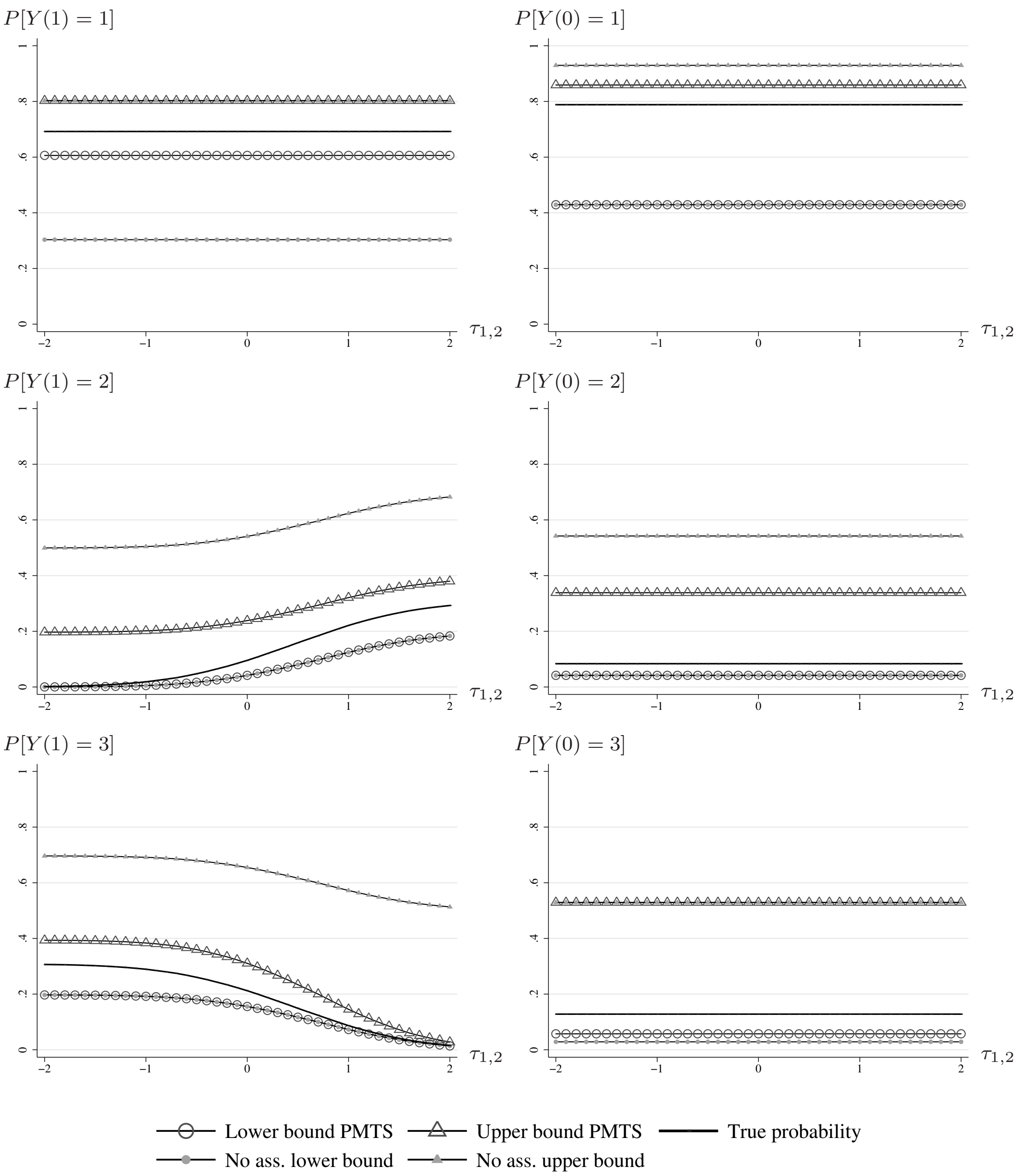
Notes: Thick black lines denote the true potential outcome probability $P[Y(s) = y]$ as indicated on the vertical axes, small grey triangles (circles) denote the no-assumptions upper (lower) bounds, large hollow triangles (circles) denote the upper (lower) bounds as derived in Proposition 1/Corollary 1 under the partial monotone treatment selection (PMTS) assumptions.

Figure 5: Bounds on Potential Outcome Distributions – Variation of $\tau_{1,1}$



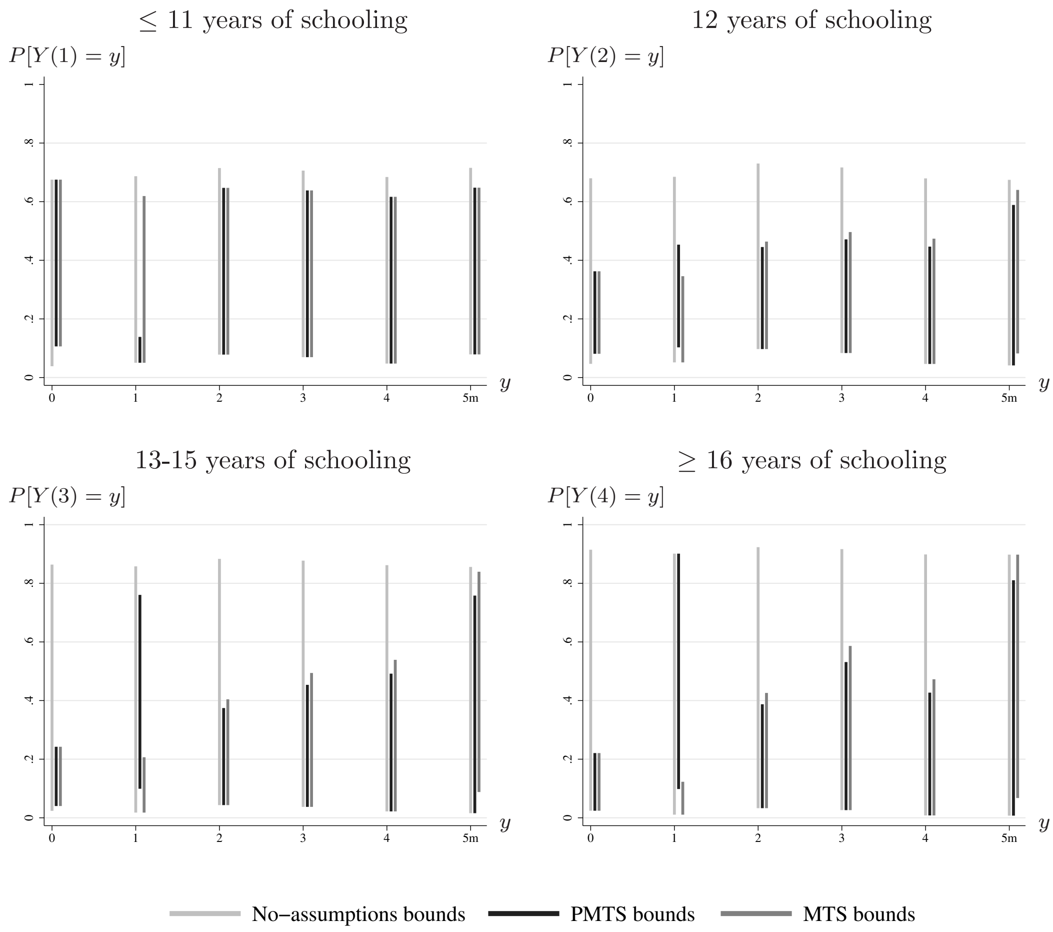
Notes: Thick black lines denote the true potential outcome probability $P[Y(s) = y]$ as indicated on the vertical axes, small grey triangles (circles) denote the no-assumptions upper (lower) bounds, large hollow triangles (circles) denote the upper (lower) bounds as derived in Proposition 1/Corollary 1 under the partial monotone treatment selection (PMTS) assumptions.

Figure 6: Bounds on Potential Outcome Distributions – Variation of $\tau_{1,2}$



Notes: Thick black lines denote the true potential outcome probability $P[Y(s) = y]$ as indicated on the vertical axes, small grey triangles (circles) denote the no-assumptions upper (lower) bounds, large hollow triangles (circles) denote the upper (lower) bounds as derived in Proposition 1/Corollary 1 under the partial monotone treatment selection (PMTS) assumptions.

Figure 7: Bounds on the Distribution of the Number of Children by Schooling Level



Source: GSS 1972-2008, own calculations. *Notes:* $P[Y(1) = y]$ denotes the potential distribution of the number of children if everybody in the population (defined as white women aged 55-70 years not in the labor force) was observed with an education level of less than 12 years. $P[Y(2) = y]$ denotes the same distribution if everybody was observed with 12 years of schooling, $P[Y(3) = y]$ with 13-15 years of schooling, and $P[Y(4) = y]$ with 16 years or more, respectively. Left bars are obtained using the empirical evidence alone, middle bars are obtained under partial monotone treatment selection (PMTS) assumptions imposing that the potential conditional transition probabilities are higher (lower) for the high educated than the low educated women for the outcome levels 0, (1), 2, 3, and 4. Outcome 5m stands for 5 or more children. Right bars are obtained under the presumption of monotone treatment selection (MTS) with the transition probabilities higher for the high eductated women over the entire support of Y.

Working Papers of the Socioeconomic Institute at the University of Zurich

The Working Papers of the Socioeconomic Institute can be downloaded from http://www.soi.uzh.ch/research/wp_en.html

0919	Rotten Kids with Bad Intentions, Nick Netzer, Armin Schmutzler, December 2009, 38 p.
0918	Partial Identification of Discrete Counterfactual Distributions with Sequential Update of Information, Stefan Boes, December 2009, 37 p.
0917	How much do journal titles tell us about the academic interest and relevance of economic reserach? An empirical analysis, Felix Schläpfer, December 2009, 14 p.
0916	Fine Tuning of Health Insurance Regulation: Unhealthy Consequences for an Individual Insurer, Johannes Schoder, Michèle Sennhauser, Peter Zweifel, August 2009, 18 p.
0915	Capping Risk Adjustment?, Patrick Eugster, Michèle Sennhauser, Peter Zweifel, September 2009, 27 p.
0914	A Pharmaceutical Innovation: Is it Worth the Money? Whose Money?, Michèle Sennhauser, Peter Zweifel, September 2009, 22 p.
0913	Copula-based bivariate binary response models, Rainer Winkelmann, August 2009, 26 p.
0912	Simulating WTP Values from Random-Coefficient Models, Maurus Rischatsch, July 2009, 6 p.
0911	Physician dispensing and the choice between generic and brand-name drugs – Do margins affect choice?, Maurus Rischatsch, Maria Trottmann, July 2009, 15 p.
0910	GPs' preferences: What price fee-for-service?, Peter Zweifel, Maurus Rischatsch, Angelika Brändle, July 2009, 21 p.
0909	Social Mobility and Preferences for Income Redistribution: Evidence from a Discrete Choice Experiment, Ilja Neustadt, Peter Zweifel, July 2009, 31 p.
0908	Robust estimation of zero-inflated count models, Kevin E. Staub, Rainer Winkelmann, June 2009, 22 p.
0907	Competitive Screening in Insurance Markets with Endogenous Wealth Heterogeneity, Nick Netzer, Florian Scheuer, April 2009, 28 p.
0906	New Flight Regimes and Exposure to Aircraft Noise: Identifying Housing Price Effects Using a Ratio-of-Ratios Approach, Stefan Boes, Stephan Nüesch, April 2009, 40 p.
0905	Patents versus Subsidies – A Laboratory Experiment, Donja Darai, Jens Großer, Nadja Trhal, March 2009, 59 p.
0904	Simple tests for exogeneity of a binary explanatory variable in count data regression models, Kevin E. Staub, February 2009, 30 p.
0903	Spurious correlation in estimation of the health production function: A note, Sule Akkoyunlu, Frank R. Lichtenberg, Boriss Siliverstovs, Peter Zweifel, February 2009, 13 p.
0902	Making Sense of Non-Binding Retail-Price Recommendations, Stefan Bühler, Dennis L. Gärtner, February 2009, 30 p.
0901	Flat-of-the-Curve Medicine – A New Perspective on the Production of Health, Johnnes Schoder, Peter Zweifel, January 2009, 35 p.
0816	Relative status and satisfaction, Stefan Boes, Kevin E. Staub, Rainer Winkelmann, December 2008, 11 p.
0815	Delay and Deservingness after Winning the Lottery, Andrew J. Oswald, Rainer Winkelmann, December 2008, 29 p.
0814	Competitive Markets without Commitment, Nick Netzer, Florian Scheuer, November 2008, 65 p.

0813 Scope of Electricity Efficiency Improvement in Switzerland until 2035, Boris Krey, October 2008, 25 p.
0812 Efficient Electricity Portfolios for the United States and Switzerland: An Investor View, Boris Krey, Peter Zweifel, October 2008, 26 p.
0811 A welfare analysis of "junk" information and spam filters; Josef Falkinger, October 2008, 33 p.
0810 Why does the amount of income redistribution differ between United States and Europe? The Janus face of Switzerland; Sule Akkoyunlu, Ilja Neustadt, Peter Zweifel, September 2008, 32 p.
0809 Promoting Renewable Electricity Generation in Imperfect Markets: Price vs. Quantity Policies; Reinhard Madlener, Weiyu Gao, Ilja Neustadt, Peter Zweifel, July 2008, 34p.
0808 Is there a U-shaped Relation between Competition and Investment? Dario Sacco, July 2008, 26p.
0807 Competition and Innovation: An Experimental Investigation, May 2008, 20 p.
0806 All-Pay Auctions with Negative Prize Externalities: Theory and Experimental Evidence, May 2008, 31 p.
0805 Between Agora and Shopping Mall, Josef Falkinger, May 2008, 31 p.
0804 Provision of Public Goods in a Federalist Country: Tiebout Competition, Fiscal Equalization, and Incentives for Efficiency in Switzerland, Philippe Widmer, Peter Zweifel, April 2008, 22 p.
0803 Stochastic Expected Utility and Prospect Theory in a Horse Race: A Finite Mixture Approach, Adrian Bruhin, March 2008, 25 p.
0802 The effect of trade openness on optimal government size under endogenous firm entry, Sandra Hanslin, March 2008, 31 p.
0801 Managed Care Konzepte und Lösungsansätze – Ein internationaler Vergleich aus schweizerischer Sicht, Johannes Schoder, Peter Zweifel, February 2008, 23 p.
0719 Why Bayes Rules: A Note on Bayesian vs. Classical Inference in Regime Switching Models, Dennis Gärtner, December 2007, 8 p.
0718 Monoplistic Screening under Learning by Doing, Dennis Gärtner, December 2007, 29 p.
0717 An analysis of the Swiss vote on the use of genetically modified crops, Felix Schläpfer, November 2007, 23 p.
0716 The relation between competition and innovation – Why is it such a mess? Armin Schmutzler, November 2007, 26 p.
0715 Contingent Valuation: A New Perspective, Felix Schläpfer, November 2007, 32 p.
0714 Competition and Innovation: An Experimental Investigation, Dario Sacco, October 2007, 36p.
0713 Hedonic Adaptation to Living Standards and the Hidden Cost of Parental Income, Stefan Boes, Kevin Staub, Rainer Winkelmann, October 2007, 18p.
0712 Competitive Politics, Simplified Heuristics, and Preferences for Public Goods, Felix Schläpfer, Marcel Schmitt, Anna Roschewitz, September 2007, 40p.
0711 Self-Reinforcing Market Dominance, Daniel Halbheer, Ernst Fehr, Lorenz Goette, Armin Schmutzler, August 2007, 34p.
0710 The Role of Landscape Amenities in Regional Development: A Survey of Migration, Regional Economic and Hedonic Pricing Studies, Fabian Waltert, Felix Schläpfer, August 2007, 34p.
0709 Nonparametric Analysis of Treatment Effects in Ordered Response Models, Stefan Boes, July 2007, 42p.